Improved and Secure Digital Watermarking for Image Tamper Detection and Localization

Elizabeth Campos-Ponce
Instituto Politecnico Nacional
SEPI ESIME Culhuacan
Mexico City, Mexico
ecamposp2000@alumno.ipn.mx

Manuel Cedillo-Hernandez
Instituto Politecnico Nacional
SEPI ESIME Culhuacan
Mexico City, Mexico
mcedilloh@ipn.mx

***Abstract*—This paper presents an improved and secure digital watermarking algorithm for image tamper detection and localization. The main challenge in every watermarking algorithm with tamper detection purposes is to create an efficient localization, robust and imperceptible scheme. Unlike conventional watermarking methods, this proposal is based on lifting wavelet transform domain in conjunction with hash-codes and key-codes criteria that conforms the hash-key codes, which are concealed into the least significant bits of each frequency component. The performance of the proposed method is evaluated in terms of imperceptibility and effectiveness to tamper detection measured by false positive, false negative and tamper detection rates, respectively. The watermarked images have high image quality in terms of peak signal to noise ratio and structural similarity index. The algorithm can handle various size of tampering, from very small to large. The proposed method has been tested for a wide variety of tampering attacks such as copy-paste, copy-move, constant average, impulsive noise, and general tampering. It can efficiently detect tampering in the presence of these attacks. Performance comparison with current state-of-the-art is included.**

Keywords—digital image watermarking, tamper detection, image processing, information security, LSB substitution

I. INTRODUCTION

Today, more and more information is being transferred electronically. Electronic data is transferred on different media, e.g., image, video and audio signals on smartphones, images via the Internet, among others. One of the problems associated with intellectual property is that if someone has access to all these data, then they have the possibility of copying, altering, and retransmitting them to unauthorized users without any restriction. This will always be possible because digital data can be reproduced identically. A promising solution to protect intellectual property and copyright information against unauthorized users is the use of the digital watermarking which generally hides certain information associated with the owner and/or distributor of e.g., an image, so that only authorized users can make legal use of the data in question [1-4]. The digital watermark may then contain information from the author, or the distributor of the image, or even the image itself, and this in turn may be visible, or invisible, depending on the application, but in both cases, it can be detected by a computer. In general, an invisible mode watermarking algorithm consists of three parts: the watermark, the encoder (insertion algorithm), and the decoder (detection algorithm). In a digital imaging context, watermarks can be applied in the spatial domain, and with the help of a transformation, they can be applied in some other domain, such as the frequency domain. It has been shown that, when the watermark is desired to be invisible to human vision, methods in the frequency domain perform better in terms of imperceptibility-robustness compared to methods based on the spatial domain [1-4].

This paper presents an improved and secure digital watermarking algorithm for image tamper detection and localization. The main challenge in every watermarking algorithm with tamper detection purposes is to create an efficient localization, robust and imperceptible scheme. Unlike conventional watermarking methods [5-10], this proposal is based on lifting wavelet transform (LWT) domain [11], in conjunction with hash-codes using RIPEMD-160 message digest algorithm [12], and key-codes using pseudo random number generator (PRNG), both conforms the hash-key codes, which are concealed using least significant bit (LSB) substitution in frequency domain. The performance of the proposed method is evaluated in terms of imperceptibility and effectiveness to tamper detection measured by false positive (FPR), false negative (FNR) and tamper detection (TDR) rates, respectively. The watermarked images have high image quality in terms of peak signal to noise ratio (PSNR) and structural similarity index (SSIM). The algorithm can handle various size of tampering, from very small to large. The proposed method has been tested for a wide variety of tampering attacks such as copy-paste, copy-move, constant average, impulsive noise, and general tampering. It can efficiently detect tampering in the presence of these attacks. Performance comparison with current state-of-the-art is included.

II. PROPOSED METHOD

This paper presents an improved and secure digital watermarking algorithm for image tamper detection and localization, which is inspired on previous work reported in [10], however, our proposal has the follow differences: A) instead of embedding the watermark data bits in spatial domain, we employ the CA band of the 1-level decomposition of LWT transform as embedding domain, B) Reduction of computation time of embedding, extraction and tampering

detection procedures, C) To improve the security, we replace the SHA-1 used in [10] with RIPEMD-160 message digest algorithm [12]. The proposed algorithm is described in the follow paragraphs. General diagram of embedding procedure is shown in Fig. 1.

A. *Embedding stage*

1) Given an image I of dimensions 512x512 in grayscale with 8 bit/pixel depth, this is divided in blocks of 8x8 pixels in size, to later apply the lifting wavelet transform (LWT) with 1-level of decomposition, obtaining the fequency bands denoted as CA, CH, CV and CD, respectively.

2) Using the CA frequency band, strip its least significant bits (LSB) and generate a 16-bit block-key.

3) Subsequently using RIPEMD-160 message digest algorithm [12], obtain the checksum and get the first 16 bits of each hash value of CA band.

4) Using the 16 bit block-key and the 16 bits of RIPEMD-160 hash-code information, employing the XOR operation, obtain the hash-key code for each CA and embedding this into its LSB using LSB subtitution. To embedding procedure, LSB substitution technique is used, this technique consists of the substitution of the pixel LSB by a bit b = {0,1}:

Considering the value of a pixel pair as $2i$, with i = 0, ..., (2^{12}/2-1) = 127, the value changes according to (1):

$$2i \xrightarrow{LSB\text{–substitution}} \begin{cases} 2i, & \text{if } b = 0 \\ 2i+1, & \text{if } b = 1 \end{cases} \quad (1)$$

Considering the value of an odd pixel as $2i + 1$, with i = 0, ..., 127, its value is given by (2):

$$2i+1 \xrightarrow{LSB\text{–substitution}} \begin{cases} 2i, & \text{if } b = 0 \\ 2i+1, & \text{if } b = 1 \end{cases} \quad (2)$$

These four steps of the embedding procedure are shown in Fig 2. Repeat this procedure of each CA frequency band. To obtain de watermarked image *I'*, apply the inverse LWT transformation using the watermarked band CA in conjunction with unaltered bands CV, CH, and CD of each 8x8 block.

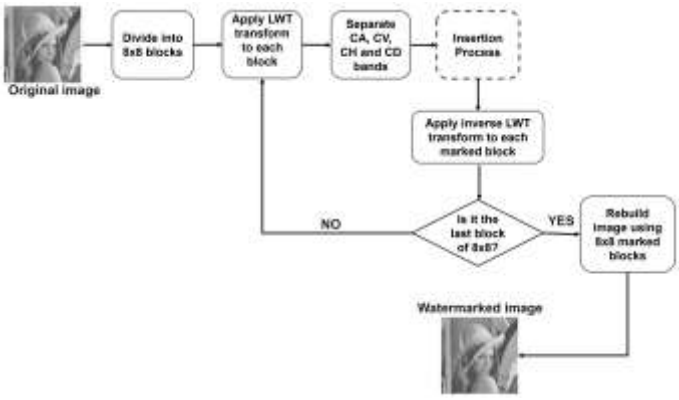


Fig. 1. General diagram of embedding procedure

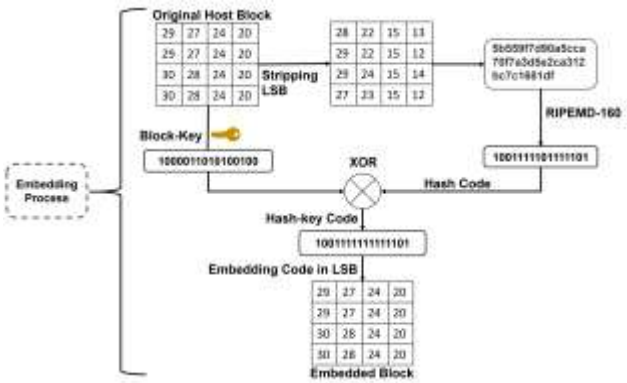


Fig. 2. Embedding procedure

B. *Extraction and tampering detection procedures*

1) Considering the watermarked image *I'*, divide this in blocks of 8x8 pixels in size, and apply the LWT transform with 1-level of decomposition, obtaining the fequency bands denoted as CA, CH, CV and CD, respectively.

2) Using the CA frequency band, strip its LSB and store these in the recovered watermark *W'*, and using CA without LSB generate a 16-bit block-key.

3) Subsequently using RIPEMD-160 message digest algorithm [12], obtain the checksum and get the first 16 bits of each hash value of CA band.

4) Using the 16 bit block-key and the 16 bits of RIPEMD-160 hash-code information, employing the XOR operation, obtain the hash-key code for each watermarked CA, the XOR result is stored in *W*.

5) To tamper detection, compute the bit error rate (BER) between *W* and *W'*, if BER=0 the 8x8 block in *Iw* is considered unaltered, otherwise a tamper is detected.

Repeat this procedure of each CA frequency band. These five steps of the extraction/tampering detection procedure procedure are shown in Fig 3.

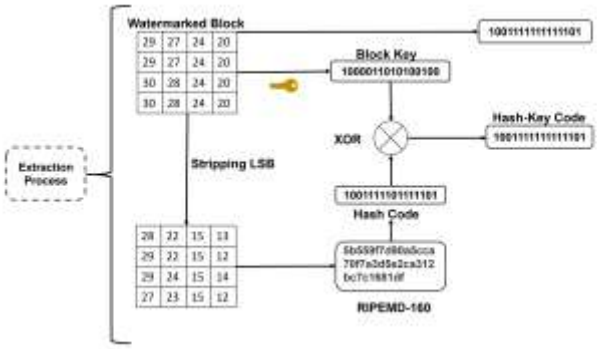


Fig. 3. Extraction procedure

III. Experimental Results and Discussion

In this section, the performance of the proposed algorithm is evaluated considering imperceptibility and tampering detection properties, using a variety of digital grayscale images. We have used 50 images with different content among which are Goldhill, Man, Lena, Airplane, Baboon, Peppers, among others, obtained from [13], all 512x512 pixels in size and grayscale resolution of 8 bits/pixel. Our experiments were carried out on a personal computer running Microsoft Windows 10© with an Intel© Core i7 processor (1.8 GHz) and 16 GB RAM while the embedding, extraction and tampering detection procedures were implemented on Matlab© 2020b.

A. Watermark imperceptibility

The watermarked image quality is measured using the following well-known indices Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) [14], given by (3) and (4) respectively.

$$PSNR(dB) = 10\log_{10}\left(\frac{255^2}{\frac{1}{N\cdot M}\left(\sum_{x=1}^{N}\sum_{y=1}^{M}\left(I(x,y)-I'(x,y)\right)^2\right)}\right) \tag{3}$$

where $N\times M$ are the original dimensions, I and I' are the original and watermarked images, respectively.

$$SSIM(I,I') = \frac{\left(2\mu_I\mu_{I'}+C_1\right)\left(2\sigma_{I'}+C_2\right)}{\left(\mu_I^2+\mu_{I'}^2+C_1\right)\left(\sigma_I^2+\sigma_{I'}^2+C_2\right)} \tag{4}$$

where C_1 and C_2 are small constant values defined in [14], I and I' are the original and watermarked images, respectively.

Considering a 1-level LWT decomposition, using the frequency bands CA, CH, CV, and CD as embedding domain each one, the average PSNR obtained is shown in Fig. 4:

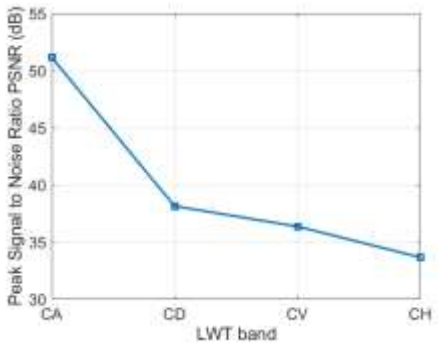


Fig. 4. Average PSNR for each frequency band of LWT transform

From Fig. 4 we show that the CA band provides the highest PSNR value. To avoid any visual distortion in the watermarked images, we adopt the CA band of LWT transform as embedding domain in the testing. Fig. 5 shows a couple of original, and its watermarked versions obtained from the method in [10] and our proposed method, with the PSNR and SSIM values obtained by each one.

Fig. 5. Watermark imperceptibility in trems of PSNR and SSIM obtained by the method in [10] and our proposed method.

B. Tampering detection

Tampered detection performance is measured using three metrics defined denoted as False Positive Rate (FPR), False Negative Rate (FNR) and Tampering Detection Rate (TDR) defined by (5), (6) and (7) respectively:

$$FPR = \frac{No.of\ \ pixels\ \ falsely\ detected\ as\ tampered}{No.of\ \ pixels\ in\ tampered\ region}\times 100 \tag{5}$$

$$FNR = \frac{No.of\ \ pixels\ \ falsely\ detected\ as\ untampered}{No.of\ \ pixels\ in\ untampered\ region}\times 100 \tag{6}$$

$$TDR = \frac{No.of\ \ detected\ tampered\ \ pixels}{Current\ tampered\ \ pixels}\times 100 \tag{7}$$

Considering the attacks of random tamper, copy-move and copy-paste with rates ranging from 1 to 50, and the USC-SIPI image data base, Fig. 6, Fig. 7, Tables I and II shows the performance obtained by the conventional method in [10] and our proposed method based on frequency domain.

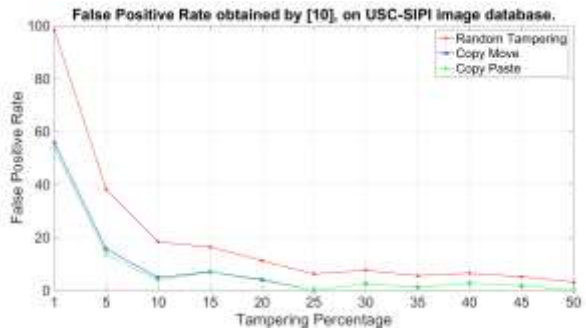


Fig. 6. Average FPR obtained by method in [10]

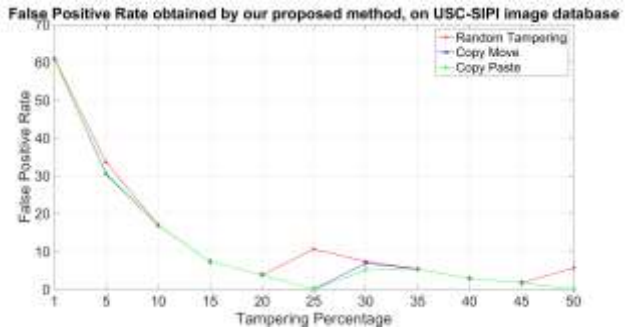


Fig. 7. Average FPR obtained by our proposed method

TABLE I. FALSE POSITIVE, FALSE NEGATIVE AND TAMPERING DETECTION RATES OBTAINED BY OUR PROPOSED METHOD, ON USC-SIPI IMAGE DATABASE.

Rate	Random tamper			Copy move			Copy paste		
	FPR	FNR	TDR	FPR	FNR	TDR	FPR	FNR	TDR
1	60.9	0	100	60.9	0	100	60.9	0	100
5	33.5	0	100	30.4	0	99	30.6	0	99
10	16.9	0	100	16.9	0	100	16.9	0	100
15	7.3	0	100	7.3	0	100	7.3	0	100
20	3.8	0	100	3.8	0	100	3.8	0	100
25	10.6	0	100	0	0	100	0	0	100
30	7.3	0	100	6.8	0	100	5.1	0	100
35	5.3	0	100	5.3	0	100	5.3	0	100
40	2.8	0	100	2.8	0	100	2.8	0	100
45	1.7	0	100	1.7	0	100	1.7	0	100
50	5.5	0	100	0	0	100	0	0	100

TABLE II. FALSE POSITIVE, FALSE NEGATIVE AND TAMPERING DETECTION RATES OBTAINED BY [10], ON USC-SIPI IMAGE DATABASE.

Rate	Random tamper			Copy move			Copy paste		
	FPR	FNR	TDR	FPR	FNR	TDR	FPR	FNR	TDR
1	98.1	0	99	54.7	0	98	53.8	0	98
5	38.2	0	99	13.6	0	99	13.6	0	99
10	18.3	0	99	3.8	0	98	3.8	0	99
15	16.3	0	98	6.7	0	99	6.8	0	99
20	11.1	0	100	3.7	0	100	3.7	0	100
25	6.1	0	100	0	0	100	0	0	100
30	7.5	0	100	2.5	0	100	2.5	0	100
35	5.4	0	100	1.1	0	100	1.1	0	100
40	6.4	0	100	2.6	0	100	2.6	0	100
45	5.1	0	100	1.7	0	100	1.7	0	100
50	3.1	0	100	0	0	100	0	0	100

From Figs. 6-7, and Tables I-II, we show that the work in [10] and our proposed method has similar performance against copy-move and copy-paste attacks, however, our proposed method outperforms to [10] when the watermarked images are attacked by random tamper, specifically when the random tamper rate is less than 5. The advantage to perform the digital watermarking in frequency instead of spatial domain is the reduction of computation time in the embedding and extraction/tampering detection procedures, as shown in Table III.

TABLE III. EMBEDDING AND EXTRACTION/TAMPERING DETECTION COMPUTATION TIMES OBTAINED BY [10] AND OUR PROPOSED METHOD

Method	Embedding time	Extraction and Tampering Detection time
[10]	5.677 s	6.766 s
Proposed Method	4.057 s	3.262 s

Performance comparison with the current state-of-the-art is shown Table IV.

TABLE IV. PERFORMANCE COMPARISON

Method	Technique	Average PSNR (dB)	Average SSIM	Comments
[5]	Spatial domain Singular Value Decomposition (SVD)	41.39	-	High computational complexity
[6]	Vector quantization	42.00	-	Visual quality affected, PSNR low
[7]	Frequency domain Faber–Schauder DWT	51.07	-	False positive rate high
[8]	Spatial domain	51.14	0.9948	Weak against copy-paste, copy-move, and constant average attacks
[9]	Spatial domain	44.50	0.9993	Visual quality affected, PSNR low
[10]	Spatial domain, hash-based embedding	51.12	0.9959	Computation time increases when the spatial resolution increases. Weak against very small random tamper.
Our	Frequency LWT domain, hash-based embedding	51.17	0.9957	Detects copy-paste, copy-move, impulsive noise and constant average attacks.

Finally, Fig. 8 shows the performance of the proposed method and algorithm in [10] against several attacks.

IV. CONCLUSIONS

At present, the information that flows through the internet is one of the main concerns of data security, so the proposed method aims to secure the information of the image owner against possible attacks or manipulations, through the watermarking technique to detect and locate possible altered

regions. The proposed approach is a frequency domain block-based image manipulation detection technique using LWT domain. Within the development of the proposed method, the RIPEMD-160 message digest algorithm was considered, and its use allows us to preserve the integrity of the content image. Finally, we know that the technique was able to detect attacks within the set of images that were used as evidence, locating the manipulations of several shapes and sizes, determining that there is some alteration in the image when applying attacks such as copy-paste, salt and pepper noise, copy-move, among others. However, there is still more opportunities to discover within the present work, such as diminish the false positive rates to random tampering, copy-move, and copy-paste attacks considering low rates.

ACKNOWLEDGMENT

Authors thank the Instituto Politecnico Nacional (IPN) as well as the Consejo Nacional de Ciencia y Tecnologia de Mexico (CONACYT) by the support provided during the realization of this research.

REFERENCES

[1] Cox I.J., Miller M.L., Bloom J.A., Fridrich J., Kalker T., "Digital watermarking and steganography", second edition, Morgan Kaufmann Publisher, San Francisco, 2009, https://doi.org/10.1016/B978-0-12-372585-1.X5001-3

[2] Barni M, Bartolini F., "Applications. In: Watermarking systems engineering: enabling digital assets security and other applications". CRC Press, Boca Raton, 2004.

[3] Lakshman Ji et.al, "Robust Digital Watermarking Techniques for protecting copyright Applied to Digital Data: A Survey", Turkish Journal of Computer and Mathematics Education, 12(3), pp. 3819-3825, 2021, https://doi.org/10.17762/turcomat.v12i3.1669

[4] P. Bas, T. Furon, F. Cayre, G. Doërr, B. Mathon, "A quick tour of watermarking techniques. In: Watermarking Security Fundamentals, Secure Design and Attacks", SpringerBriefs in Electrical and Computer Engineering. Springer, Singapore, pp. 13–31, 2016.

[5] Qi X, Xin X., "A singular-value-based semi-fragile watermarking scheme for image content authentication with tamper localization", J Vis Commun Image Represent 30:312–327, 2015, https://doi.org/10.1016/j.jvcir.2015.05.006

[6] Tiwari A, Sharma M, Tamrakar R. K., "Watermarking based image authentication and tamper detection algorithm using vector quantization approach", AEU Int J Electron Commun 78:114–123, 2017. https://doi.org/10.1016/j.aeue.2017.05.027

[7] Azeroual A., Afdel K., "Real-time image tamper localization based on fragile watermarking and Faber-Schauder wavelet", AEU Int J Electron Commun 79:207–218, 2017, https://doi.org/10.1016/j.aeue.2017.06.001

[8] Gull S, Loan NA, Parah SA et al. "An efficient watermarking technique for tamper detection and localization of medical images", J Amb Intell Humaniz Comput pp.1–10, 2018, https://doi.org/10.1007/s12652-018-1158-8

[9] Sarkar D., Palit S., Som S., Dey K.N., "Large scale image tamper detection and restoration", Multimed Tools Appl., 2020, https://doi.org/10.1007/s1104 2-020-08669 -0

[10] Bhalerao, S., Ansari, I.A. & Kumar, A., "A secure image watermarking for tamper detection and localization", J Ambient Intell Human Comput 12, 1057–1068, 2021, https://doi.org/10.1007/s12652-020-02135-3

[11] Sweldens, W., "The Lifting Scheme: a Construction of Second Generation of Wavelets," SIAM J. Math. Anal., 29 (2), pp. 511–546, 1998.

[12] C. Paar and J. Pelzl, "Hash Functions," in Understanding Cryptography: A Textbook for Students and Practitioners, Heidelberg, Berlin, Germany: Springer-Verlag, 2010, pp. 293-317.

[13] USC SIPI Image Database available on: http://sipi.usc.edu/database/SIPI_Database.pdf

[14] Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P., "Image quality assessment: from error measurement to structural similarity", IEEE Transactions on Image Processing, 2004, vol. 13, no. 4, p. 600–612. https://doi.org/10.1109/tip.2003.819861

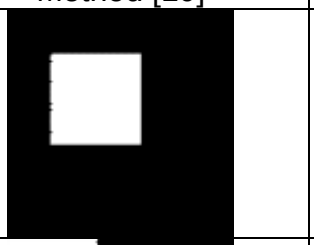
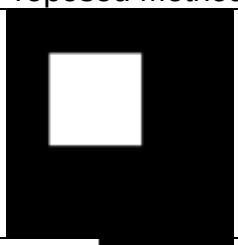
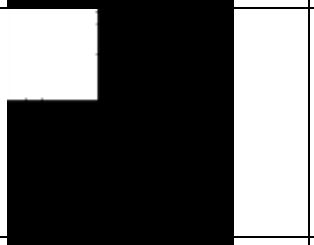
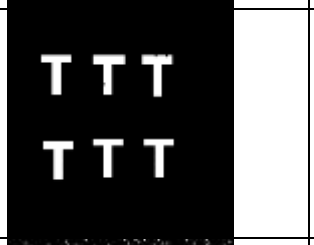
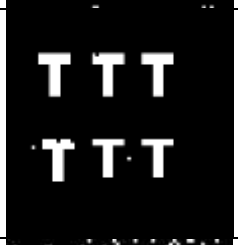
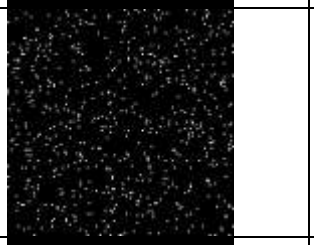
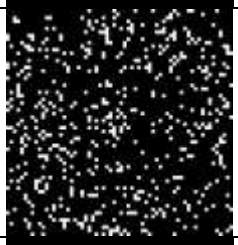
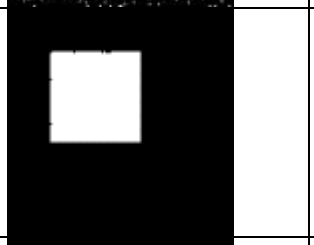
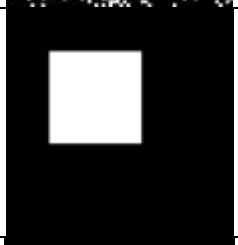
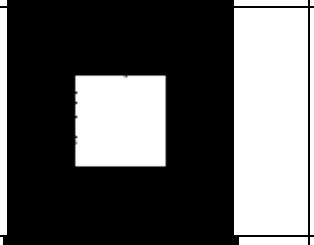
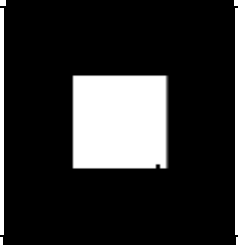
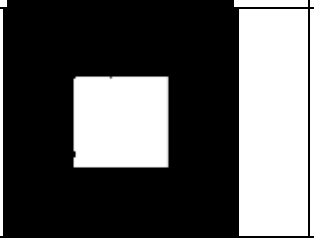
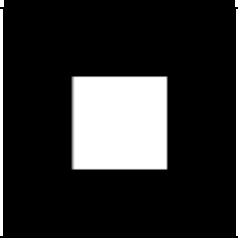
Attack	Distorted Image	Tampered Mask of Method [10]	Tampered Mask of Proposed Method
Copy-Paste			
Copy-Move			
Text addition			
Impulsive noise with density of 0.005			
Constant Average			
Central crop (on)			
Central crop (off)			

Fig. 8. Performance of the watermarking algorithm of [10] and our proposed method.